

Use of Cluster Analysis in Exploring Economic Indicator Differences among Regions: The Case of Latvia

Ieva Brauksa

Abstract—This paper can help decision makers to identify the regions with the largest need for stimulating their development. It offers ideas how to compare municipalities and how to search for common factors that influence development. Data from Latvia are used as an example of this method. It shows how cluster analysis can be applied for comparing socio-economic development of different municipalities and confirms the hypothesis that one of the characteristics common for municipalities with similar measures describing development is their geographical location.

Index Terms—Cluster analysis, municipalities, regional development.

I. INTRODUCTION

Though this paper is focused on the example of Latvia, it can be useful also for researchers from other countries as it offers ideas how to group and compare municipalities and evaluate efficiency of territorial reforms. It gives a new perspective on how to identify regions with the largest need for stimulating their development. Main suggestion is to group municipalities based on criteria that characterize their social and economic development (selection of the variables included in analysis should be careful and strictly based on the estimates that is believed to matter most for development in particular country).

The next step after the municipalities are grouped based on similarities in their development indicators, further analysis is needed to check what other features are common for these municipalities. The proposed hypothesis is that one of the characteristics that will be common for municipalities in the same group is their geographical location, i.e. municipalities from the same region will tend to be part of the same groups. In case if this hypothesis is fulfilled, country experience regional stratification and support should be allocated not just for particular municipalities, but planned in regional level.

This paper will present summary of previous research about region comparison using cluster analysis and describe methodology that can be used for this task. The elaborated example based on data from Latvia is included to show how cluster analysis can be applied in municipality comparison. After administrative territorial reform in 2009 Latvia was split in 109 municipalities. There have been wide debates

about borders and differences of these municipalities. In this paper analysis will be used to look at the question of differences among municipalities from another perspective. It uses some variables that describe social and economic development base for municipalities (mainly focusing on the potential development based on labour force availability and entrepreneurship activities) and checks what kind of municipalities have similar results.

II. PREVIOUS RESEARCH

Previously authors have used various methods for municipality comparison and evaluation of regional development. As the focus of this paper is to explore possibilities to apply cluster analysis to solve this question, previous research analysis will mainly deal with authors that have used this tool for region comparison in different countries.

Cluster analysis is one of multidimensional methods of statistical analysis that allows classifying observations in groups. Observations with similar attributes are grouped together therefore variables for this analysis should be chosen with care as this choice can lead to different results. Cluster analysis is often derogated as it has different methods and each of them can have different results, but in the case of municipality analysis the aim is to find out the main common lines and slightly different municipality classification in one or another group is not so essential.

Previous research has used cluster analysis for different set of countries and various areas of regional analysis. For example, Stimson et al. [1] use cluster analysis for ten years of data for cities in Australia offering to use the results of this research to encourage stronger development of less developed regions, but Petterson [2] by this type of analysis studies data on Sweden revealing significant differences in socio-economic conditions between populations living in different places.

There is also a discussion of variable which should be included in analysis. Of course, this choice depends also on data availability in different countries, but some general rules still can be made. Leschke [3] notes that countries are in most cases clustered on the bases of different general characteristics of labour market and social policy. Author notes that cluster analysis has more disadvantages when there are larger variations in institutional peculiarities among countries. It is indicated that variables for the cluster analysis should be chosen such that are close to the question of interest.

Cluster analysis is often used also because of its advantage to represent data also in graphical way as it offers several

Manuscript received December 15, 2012; revised February 10, 2013. This work was supported by the European Social Fund within the project "Support for Doctoral Studies at University of Latvia".

I. Brauksa is with University of Latvia, Aspazijas blvd. 5, Riga, LV 1050, Latvia (e-mail: ieva.brauksa@inbox.lv).

possibilities to do that. Dendrogram or other graphical solutions are often used to visualize linkages between observations. For example, Galic et al. [4] have analyzed Bosnia and Herzegovina data and provides data visualization exemplified. Cluster analysis in this case is used to group cities with similar social economic development, though in this case authors have chosen to group them only based on growth rate of GDP per capita and illiteracy per capita and share of population employment in tertiary sector.

An important question of interest that most often stimulate these researches is evaluation of regional development both in rural-urban and in comparative context. Vincze & Mezei [5] demonstrate the role of cluster analysis of rural localities in the process of choosing the rural development measures to be used to stimulate rural socio-economic growth. They perform two stage analysis using large number of variables grouped by factor analysis and later in cluster analysis. Authors note that these methods cannot give a general solution because of the problems of abstractions, but this information could represent a good basis for a detailed analysis.

Similarly also Mistre & Muska [6] as part of their research to compare economic development levels across districts of Latvia used cluster analysis using a list of statistical indicators for the period till the administrative territorial reform in 2009. They confirm that in Latvia there is a monocentric economic development trend and therefore there are significant differences between capital city and other districts.

Another context of this kind of analysis is not just regional comparison but search for the evidence and reasons of inequality. This inequality could be measured by the level of income, unemployment of other measures describing social and economic situation in regions. For example, Peters [7] identifies and describes clusters of persistently low and high income inequality in the North Central Region.

Using more macro level data Jurun & Pivac [8] uses cluster analysis during their studies of regional GDP of Croatian counties. These counties are classified using cluster analysis to make a comparative analysis with official region division which is more geographical and political areas than real homogenous socio-economic areas. Similarly also Poder et al. [9] have analysed Estonian data to study entrepreneurial activity in rural municipalities by applying hierarchical cluster analysis.

Therefore previous research analysis shows that cluster analysis is already widely applied in analysis different topics of regional development both based on micro and macro level data. Separate studies have used examples from both large and small countries and compared situation within one country and across some larger regions.

This paper is aimed to contribute to this literature with the analysis of Latvian data which is specific as there recently has been administrative territorial reform in this country and further analysis is needed to see if municipalities that are located nearby have similar economic indicators. Moreover this paper will show how cluster analysis can be used in identifying if some economic development problems exist only in particular municipalities or the problem is wider as well as there is a search of evidence if higher development in

nearby municipalities can promote development in wider area.

III. METHODOLOGY AND DATA USED

Data from the reports of the Ministry of Regional development and local governments in Latvia will be used [10]. Cluster analysis will be based on information about unemployment rates, age structure (proportion of inhabitants at working-age), rate of change in number of permanent inhabitants and number of individual entrepreneurs in these municipalities.

These variables are chosen to reflect economic situation in municipalities and they are comparable also for regions of different size and population. Data about the changes in the number of inhabitants compares changes over period of five years (2007-2012). Data about number of inhabitants at working age and unemployment level is for the beginning of 2012, but number of individual entrepreneurs per 1000 inhabitants is used from year 2010 as it is the latest statistics available.

Cluster analysis can be a useful tool for arranging observations into groups that can be further analyzed, it is a process of finding out unknown group profiles. There have been discussions that, although now municipalities are larger than previous districts and it could help distributing resources equally and enhance equivalent development, in reality also the new system has faced several problems.

The aim of this paper is to use cluster analysis for creating new perspective for discussing similarities and differences of economic and social development in municipalities of Latvia. Later similar ideas could be applied also in the context of other countries.

Cluster analysis can be rather subjective as results may depend on method chosen. That is why in this paper several methods will be tested and compared to understand if there are significant differences in results or they show similar tendencies. K-means clustering is used to split observations in groups and test what are the main characteristics of these groups.

Cluster analysis will be used to give possibility to visually evaluate, which of the regions are more similar. More detailed statistical measures and indicators will be included to interpret the results based on numeric data. This analysis gives us a possibility to understand, if clusters, that will be formed based on this analysis, will group together municipalities from the same planning regions (Riga, Vidzeme, Kurzeme, Zemgale, Latgale) or they will be mixed differently.

This research can help to better understand situation of differences between municipalities and generate ideas what factors should be taken into account when discussing about sustainable and equal development of different regions in Latvia.

IV. RESULTS OF MUNICIPALITY COMPARISON

As there are four planning regions in Latvia (excluding capital city), analysis will be started by applying K-means

clustering method dividing all municipalities in four groups to test a simple hypothesis if municipalities from the same planning region will tend to be in a single cluster.

Administrative territorial reform was implemented in Latvia in 2009 and it divided country in 109 municipalities (after some modification later this number changed slightly), previously there were more than 500 smaller municipalities. Therefore long term analysis us territories is encumbered as the territory structure has changed.

When joining small municipalities and shifting borders there was an idea that larger territories will promote faster and economic development as larger regional units are more efficient in attracting financial support and some inequality between parts of the territories could be lessened. Though, in case if economic development is problematic not just in a single municipality but in wider region, possibly that also territorial reform is not a universal cure for this problem. That is why this paper tests similarity between municipalities based on economic criteria and checks if problematic and more advanced municipalities are from the same or from different regions.

Analysis excluded large cities, but used data only at municipality level. Table I shows the results from such analysis. Firstly, there is a specific cluster (cluster 3) with just one observation for a specific municipality Marupe which is situated next to capital city and have significantly higher rate of number of individual entrepreneurs per 1000 inhabitants as well as it had high inflow of permanent inhabitants and also the unemployment level in this municipality is much lower than in all other clusters. This municipality is very different from all other showing the highest development potential.

Territory development index [10] for 2011 and also few previous years shows that Marupe municipality was the second most developed municipality in Latvia. Though, alternative territory development index offered by Brauksa [11] suggests that traditional territory development index for this municipality is underestimated and it has the strongest development potential. Also the results of cluster analysis allow having similar conclusion and confirms this idea once more.

TABLE I: FINAL CLUSTER CENTRES

	Cluster			
	1	2	3	4
Inhabitant changes, %	17.6	-6.9	40.1	-1.9
Inhabitants working age, share	65.9	65.2	65.2	66.1
Unemployment level	5.3	13.1	3.9	8.2
Number of individual entrepreneurs per 1000 inhabitants	36.8	11.9	63.8	21.8
Number of municipalities in cluster	8	61	1	40

Opposite situation is observed in cluster 2 which has the lowest rate of individual entrepreneurs per 1000 inhabitants, highest level of unemployment as well as fastest decrease of number of permanent inhabitants over last five years. This cluster is the largest and includes 61 municipalities mainly

from Latgale (31% of them), we should note that all Latgale municipalities are included in this cluster. There are also municipalities from Vidzeme (25%), Zemgale (23%) and Kurzeme planning regions (16%), only 5% of them are from Riga planning region.

Cluster 1 is with second highest rate of increase in number of permanent inhabitants supported by relatively low unemployment rate and average high level of inhabitants involved in individual entrepreneurship. This cluster that has the second "best" measurements includes municipalities only from Riga planning region indicating that situation in municipalities near Riga is better than elsewhere in Latvia.

Cluster 4 features moderate decrease in number of permanent inhabitants (1.9% per last five years), but has the largest share of inhabitants at working age. Municipalities in this cluster on average had a bit higher unemployment rate than national average and the number of individual entrepreneurs per 1000 inhabitant was smaller than in Cluster 1 and Cluster 4 but higher than in Cluster 2. Cluster 4 includes municipalities from Riga planning region (26%), Vidzeme (16%), Kurzeme (13%) and Zemgale (10%).

Therefore we can conclude that hypothesis that even though the grouping was based on economic variables, municipalities located geographically close together also group together is confirmed. That means that situation with high unemployment rate, decrease in number of inhabitants, low share of inhabitants at working age and low entrepreneurship activity is problem not just for particular municipality, but for wider regions.

Looking at these results from another perspective we can say that also economic development encourage stronger further development. Recently there has been wide discussion about economic and development clusters meaning some territories or industries collaborating to achieve faster economic growth.

V. CONCLUSION

This paper has showed how analysis of social and economic indicators can be used to search for similarities across municipalities and get better understanding about situation in region in whole country. In case there are mechanisms for support for individual municipalities but analysis shows that similar lagged development is for several municipalities from the same region, support mechanism should be reconsidered and collaboration of smaller municipalities should be encouraged.

For the case of Latvia cluster analysis based on inhabitant structure and basic economic indicator analysis shows that there are some regional similarities for municipalities of Pieriga and Latgale. Other planning regions do not create separate groups. There are similar municipalities across all of them.

To sum up, there are different tools for exploring differences in economic development in municipalities, and cluster analysis also proves to be useful way to look at these questions from another perspective, though of course decisions and their implication should be based on

consideration of wide range of data and analysis, but this could be a part of it.

REFERENCES

- [1] R. Stimson, S. Baum, and K. O'Connor, "The Social and Economic Performance of Australia's Large Regional Cities and Towns: Implications for Rural and Regional policy," *Australian Geographical Studies*, vol. 41, no. 2, pp. 131-147, July 2003.
- [2] O. Pettersson, "Microregional fragmentation in a Swedish county," *RSAI*, vol. 80, pp. 389-409, 2001.
- [3] J. Leschke, "Is it useful to cluster countries? Analysis on the example of unemployment insurance coverage of non-standard employed," revised in *TLM.NET Conference Paper*, Budapest, 2005.
- [4] M. Galic, M. El-Zayat, J. Sero, E. Kirlic, A. Djonko, and J. Osmankovic, "Visualizing region development differences," *Conference Proceedings: International Conference of the Faculty of Economics Sarajevo (ICES)*, pp. 1-9, 2010.
- [5] M. Vincze and E. Mezei, "The increase of rural development measures efficiency at the micro-regions level by cluster analysis. A Romanian case study," *Eastern Journal of European Studies*, vol. 2 issue 1, pp. 13-39, June 2011.
- [6] B. Mistre and A. Muska, "Synergy of the amount of social security benefits and economic activity in Latvia," *European Integration Studies*, no. 5, pp. 63-71, 2011.
- [7] D. J. Peters, "Place-Based Income Inequality Clusters in the Rural North Central Region, 1979-2009," *Applied Economic Perspectives and Policy*, vol. 33, no. 2, pp. 222-240, 2011.
- [8] E. Jurun and S. Pivac, "Comparative regional GDP analysis: case study of Croatia," *CEJOR*, vol. 19, pp. 319-335, 2011.
- [9] A. Poder, M. Nurmet, and R. Varnik, "Entrepreneurial Activity in Rural Municipalities of Three South-Eastern Estonian Counties in 2005-2009: a Cluster Analysis of Rural Municipalities," *Economic Science for Rural Development Conference Proceedings*, issue 26, pp. 182-191, 2011.
- [10] State Regional Development Agency of Latvia. Data on regional development index and regional statistics. [Online]. Available: <http://www.vraa.gov.lv>
- [11] I. Brauksa, "Territory Development Index – for some territories overestimated, for some – underestimated?" *RTU scientific journal*, Riga, Latvia, 2012.



Ieva Brauksa holds a Master's Degree in Economics (with honours) from the University of Latvia (2010), previously graduating the same university as Bachelor in Economics in 2007. Currently she is studying in a doctorate program in University of Latvia in the field of Economics (Econometrics) and her current research interest is regional development and political

factor influence on economics.

Since 2011 she works as econometrician in the finance sector. Previously she has worked as consumer data analyst and researcher dealing with market research questions and as an information consultant in the agency informing society about the European Union.

Comments and questions are welcomed, you can contact the author at ieva.brauksa@inbox.lv.